US006173293B1

(12) United States Patent
Thekkath et al.

(10) **Patent No.:** **US 6,173,293 B1**
(45) **Date of Patent:** **Jan. 9, 2001**

(54) **SCALABLE DISTRIBUTED FILE SYSTEM**

(75) Inventors: **Chandramohan A. Thekkath**, Palo Alto; **Timothy P. Mann**, Redwood City; **Edward K. Lee**, Mountain View, all of CA (US)

(73) Assignee: **Digital Equipment Corporation**, Maynard, MA (US)

(*) Notice: Under 35 U.S.C. 154(b), the term of this patent shall be extended for 0 days.

(21) Appl. No.: **09/039,057**

(22) Filed: **Mar. 13, 1998**

(51) **Int. Cl.**^{7} **G06F 17/30**
(52) **U.S. Cl.** **707/201**
(58) **Field of Search** 707/1-10, 100-104, 707/200-206

(56) **References Cited**

U.S. PATENT DOCUMENTS

5,465,365	*	11/1995	Winterbottom	707/101
5,623,666	*	4/1997	Pike et al.	707/200
5,740,367	*	4/1998	Spilo	709/201
5,764,972	*	6/1998	Crouse et al.	707/1

OTHER PUBLICATIONS

Anderson et al., "Serverless Network File Systems," ACM Transactions on Computer Systems, vol. 14, No. 1, Feb. 1996, p. 41–79.
Birrell et al., "A Universal File Server," IEEE Transactions on Software Engineering, vol. SE–6, No. 5, Sep. 1980.
Devarakonda et al., "Recovery in the Calypso File System," ACM Transactions on Computer Systems, vol. 14, No. 3, Aug. 1996, pp. 287–310.
Howard et al., "Scale and Performance in a Distributed File System," ACM Transactions on Computer Systems, vol. 6, No. 1, Feb. 1988, pp. 51–81.
Johnson et al., "Overview of the Spiralog File System," Digital Technical Journal, vol. 8, No. 2, 1996.
Kazar et al., "DEcorum File System Architectural Overview," USENIX Summer Conference, Jun. 11–15, 1990, Anaheim, California.
Kronenberg et al., "VAXclusters: A Closely–Coupled Distributed System," ACM Transactions on Computer Systems, vol. 4, No. 2, May 1986, pp. 130–146.
Mann et al., "A Coherent Distributed File Cache with Directory Write–Behind," ACM Transactions on Computer Systems, vol. 12, No. 2, May 1994, pp. 123–164.
Sandberg et al., "Design and Implementation of the Sun Network Filesystem," Sun Microsystems Inc., Mountain View, California.
Shillner et al., "Simplifying Distributed File Systems Using a Shared Logical Disk," Dept. of Computer Science, Princeton University.

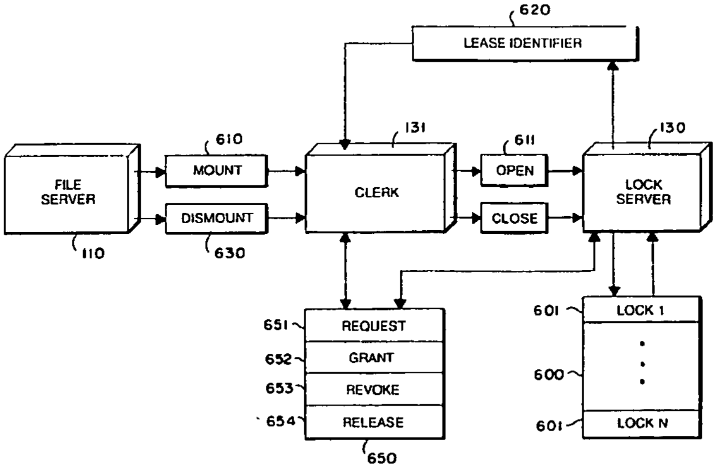
* cited by examiner

Primary Examiner—Ruay Lian Ho
(74) *Attorney, Agent, or Firm*—Pennie & Edmonds LLP

(57) **ABSTRACT**

A file system is distributed over a plurality of computers connected to each other by a network. The plurality of computers execute user programs, and the user programs access files stored on a plurality of physical disks connected to the plurality of computers. The file system includes a plurality of file servers executing on the plurality of computers as a single distributed file server layer. In addition, the file system includes a plurality of disk servers executing on the plurality of computers as a single distributed disk server layer, and a plurality of lock servers executing on the plurality of computers as a single distributed lock server to coordinate the operation of the distributed file and disk server layers so that the user programs can coherently access the files on the plurality of physical disks. The plurality of file servers executes independently on a different one of the plurality of computers, and the plurality of file servers communicate only with plurality of disk servers and the plurality of lock servers, and not with each other. Furthermore, the disk server layer organizes the plurality of physical disks as a single virtual disk having a single address space for the files.

20 Claims, 6 Drawing Sheets


620
LEASE IDENTIFIER
610
131
611
130
FILE SERVER
MOUNT
DISMOUNT
CLERK
OPEN
CLOSE
LOCK SERVER
110
630
651
REQUEST
652
GRANT
653
REVOKE
654
RELEASE
650
601
LOCK 1
600
601
LOCK N